# SPARSE RANDOM MATRICES AND STATISTICS OF ROOTED TREES

A.Khorunzhy*
Laboratoire de Physique Mathématique et Géometrie
Université Paris 7 - Denis Diderot, FRANCE†

February 1, 2008

**Abstract**

We consider the ensemble of $N \times N$ random symmetric matrices $\{A_{N,p}\}$ with independent entries such that $A_{N,p}$ has, in average, $p$ non-zero elements per row. We study the asymptotic behaviour of the maximal (in magnitude) eigenvalue $\lambda_{\max}(N)$ of matrices $A_{N,p}/\sqrt{p}$ when they are large and sparse, i.e. in the limit $N, p \to \infty$, $p = o(N)$. We prove that the value $p_c = \log N$ is the critical one for $\lambda_{\max}(\infty)$ to be bounded.

Our arguments are based on the calculus of the tree-type graphs for random matrices with independent entries. Asymptotic properties of sparse random matrices essentially depend on the typical degree of a tree vertex that we prove to be finite.

*AMS Classification numbers (1991):* 15A52 (Primary) 05C30, 60F99 (secondary)

*Key words:* random matrices; eigenvalue distribution; enumeration of graphs; Erdös-Rényi partial sums.

## 1 Introduction

The study of eigenvalue distribution of $N$-dimensional random matrices in the limit $N \to \infty$ has been initiated by E.Wigner [14]. He considered the ensemble of real symmetric matrices

$$A_N(x, y) = a(x, y), \quad x, y = 1, \ldots, N \tag{1.1}$$

*The author thanks the Department of Mathematics and Statistics of the Brunel University of West London, where this work has been started. The financial support of the Royal Society (London) is gratefully acknowledged.

†On leave of absence from Mathematical Division, Institute for Low Temperature Physics, Kharkov, UKRAINE

where $\{a(x,y), x \le y,\ x,y \in \mathbf{N}\}$ are jointly independent random variables with zero mean and variance 1. The celebrated semicircle (or Wigner) law states that the distribution function of eigenvalues $\lambda_1^{(N)} \le \ldots \le \lambda_N^{(N)}$ of the matrix $\hat{A}_N = A_N/\sqrt{N}$ given by relation

$$\sigma(\lambda; \hat{A}_N) = \#\{j : \lambda_j^{(N)} \le \lambda\} N^{-1} \tag{1.2}$$

converges in the average as $N \to \infty$ to the nonrandom distribution $\sigma_w(\lambda)$

$$\sigma_w'(\lambda) = \begin{cases} (2\pi)^{-1}\sqrt{4-\lambda^2}, & \text{if } |\lambda| \le 2 \\ 0, & \text{if } |\lambda| > 2 \end{cases} \tag{1.3}$$

provided all even moments of $a(x,y)$ exist and the odd ones vanish.

Wigner introduced random matrices (1.1) to model distribution of energy levels of heavy atomic nuclei. Such a nucleus consists of large number ($N \sim 100$) of particles that interact each with other. According to this, the statistical description naturally presumes consideration of random matrices whose entries are random variables of the same order of magnitude (see e.g. [11]).

Recent applications of random matrices lead to various generalizations of the Wigner ensemble $\{\hat{A}_N\}$. In particular, in the neural network theory the dilute version of (1.1) is used to describe the system, where some randomly chosen links between elements are broken (see e.g. [1]). Such a dilute matrix can be determined as

$$A_{N,d}(x,y) = a(x,y) d_N(x,y), \quad d_N(x,y) = d_N(y,x),$$

where $\Lambda_N \equiv \{d_N(x,y),\ x \le y,\ x,y = 1,\ldots,N\}$ is a family of independent random variables, also independent from $\{a(x,y)\}$, with the distribution

$$d_N(x,y) = \begin{cases} 1, & \text{with probability } p/N\ ; \\ 0, & \text{with probability } 1 - p/N; \end{cases} \quad p \le N. \tag{1.4}$$

It is not hard to see that the $A_{N,d}$ has, in average, $p$ non-zero elements per row. Thus, $\{A_{N,d}\}$ could be regarded as the ensemble of matrices intermediate between the discrete version of random Schroedinger operator and the Wigner random matrices. This explains the interest to dilute random matrices from the spectral theory point of view (see e.g. [9]).

The asymptotic properties of the eigenvalue distribution of $A_N^{(d)}$ were considered in different aspects [7, 10, 12]. In particular, it was shown in [12] and proved in [9] that the eigenvalue distribution of the matrix

$$A_N^{(p)}(x,y) = a(x,y)\hat{d}_N(x,y), \quad \hat{d}_N(x,y) = \frac{1}{\sqrt{p}} d_N(x,y) \tag{1.5}$$

converges in probability in the limit $N, p \to \infty$ to the semicircle distribution $\sigma_w(\lambda)$ (1.3);

$$\text{p-}\lim_{N,p\to\infty} \sigma(\lambda; A_N^{(p)}) = \sigma_w(\lambda) \tag{1.6}$$

that means that the semicircle law holds for dilute random matrices.

The eigenvalue distribution function concerns a fraction of eigenvalues of the matrix. In applications it is often important to know whether the norm of $A_N^{(p)}$ is bounded (see e.g. [5]), i.e. whether there are eigenvalues of $A_N^{(p)}$ outside of the support of $\sigma_w'(\lambda)$ in the limit $N, p \to \infty$.

In present paper we consider the probability distribution of the spectral norm

$$\|A_N^{(p)}\| = \max\{|\lambda_1^{(N,p)}|, |\lambda_N^{(N,p)}|\} \equiv \lambda_{\max}^{(N,p)},$$

where $\lambda_1^{(N,p)} \leq \ldots \leq \lambda_N^{(N,p)}$ are eigenvalues of $A_N^{(p)}$. We study the asymptotic behaviour of $\|A_N^{(p)}\|$ when matrices $A_N^{(p)}$ are large and sparse, i.e. in the limit $N, p \to \infty$, $p = o(N)$. Our results show that the rate $p_c = \log N$ is the critical one for the limit of $\|A_N^{(p)}\|$ to be either bounded (in this case it is equal to 2) or unbounded.

## 2 Main results and scheme of the proof

First of all, let us note that one can determine the collection of random variables $\Lambda \equiv \{\Lambda_N, N \in \mathbf{N}\}$ on the same probability space $\Omega_d$ with the help of the procedure due to C.Newman (see e.g. [4]). We also determine $\{a(x,y), x, y \in \mathbf{N}\}$ on the same $\Omega_a$.

We denote the measure and mathematical expectation corresponding to $\Lambda$ by $\mu_d$ and angles $\langle \cdot \rangle$, respectively. We also denote by $\mathbf{E}\{\cdot\}$ the mathematical expectation with respect to the measure $\mu_a$ generated by the random variables $\{a(x,y), x, y \in \mathbf{N}\}$.

We assume that the random variables $a(x,y)$ satisfy conditions:

$$\mathbf{E}[a(x,y)]^{2k+1} = 0 \quad \forall x, y, k \in \mathbf{N}, \tag{2.1$a$}$$

and

$$\mathbf{E}[a(x,y)]^2 = 1, \quad \sup_{x,y} \mathbf{E}[a(x,y)]^{2k} = V_{2k} < \infty \quad \forall k \in \mathbf{N}. \tag{2.1$b$}$$

**Theorem 2.1**

*Assume that the random variables $a(x,y)$ are such that there exists $\delta > 0$ such that*

$$V_{2k} \leq k^{\delta k} \quad \forall k \in \mathbf{N}. \tag{2.2}$$

*Then in the limit $N, p \to \infty$ such that $p$ satisfies condition*

$$\frac{p}{(\log N)^{1+\delta'}} \to \infty$$

*with $\delta' > 2\delta$ the norm $\|A_N^{(p)}\|$ is bounded;*

$$p\text{-}\lim_{N,p\to\infty} \|A_N^{(p)}\| = 2. \tag{2.3}$$

*Remarks.*

1. In fact, we prove that under conditions of theorem 2.1 the estimate

$$\mu_a \otimes \mu_d\{\omega : \|A_N^{(p)}\| > 2(1+2\varepsilon)\} \le \exp\{-\varphi_N(\varepsilon,\delta)\log N\} \tag{2.4}$$

holds with $\varphi_N(\varepsilon\delta) = O((\log N)^{\delta})$. This implies boundedness of $\lim \|A_N^{(p)}\|$ with probability 1. This fact together with (1.6) implies (2.3).

2. If there exists such $U$ that $V_{2k} \le U^{2k}$, then (2.4) and (2.3) hold for all $\delta' > 0$.

To show that condition $p \gg (\log N)^{1+\delta}$ is necessary for convergence (2.3), we consider the simplest case of Bernoulli random variables

$$\hat{a}(x,y) = \begin{cases} 1, & \text{with probability } 1/2; \\ -1, & \text{with probability } 1/2. \end{cases}$$

**Theorem 2.2**

*Let $N, p$ increase infinitely in the way that*

$$\frac{p}{(\log N)^{1-\delta}} \to 0 \tag{2.5}$$

*for any positive $\delta > 1$. Then for any given $R > 0$*

$$\lim_{N,p\to\infty} \mu_a \otimes \mu_d\{\omega : \|\hat{A}_N^{(p)}\| > R\} = 1, \tag{2.6}$$

*where $\hat{A}_N^{(p)}$ is given by (1.4), (1.5) with $a(x,y)$ replaced by $\hat{a}(x,y)$.*

Now we describe the scheme of the proof of Theorem 2.1. First of all, let us recall that the semicircle law proved in [14] states that the moments

$$M_j^{(N)} \equiv \mathbf{E}\int \lambda^j \mathrm{d}\sigma(\lambda;\hat{A}_N) = \mathbf{E}\frac{1}{N}\ \mathrm{Tr}\ \hat{A}_N^j \tag{2.7}$$

converges in the limit $N \to \infty$ to the numbers

$$m_j = \int \lambda^j \mathrm{d}\sigma_w(\lambda), \quad m_j = \begin{cases} t_k, & \text{for } j = 2k; \\ 0, & \text{for } j = 2k+1, \end{cases},$$

where

$$t_k = \frac{(2k)!}{k!(k+1)!}. \tag{2.8}$$

Regarding the average

$$\mathbf{E}\frac{1}{N}\ \mathrm{Tr}\ \hat{A}_N^j = \frac{1}{N}\sum_{x=1}^{N}\frac{1}{N^{j/2}}\sum_{y_i=1}^{N}\mathbf{E}a(x,y_1)a(y_1,y_2)\cdots a(y_{j-1},x) \tag{2.9}$$

as a sum over paths $W(x,Y_{j-1}) \equiv (y_0 = x, y_1, y_2, \dots, y_{j-1}, y_0)$ of $j$ steps, Wigner observed that the leading contribution to (2.7) is given by the paths that are related with the simple walks in the upper half-plane started and ended at zero. Basing on this fact, he derived recurrent relations for the limiting moments

$$t_k = \sum_{l=0}^{k-1} t_{k-1-l}\, t_l, \quad t_0 = 1. \tag{2.10}$$

In paper [8] the paths $W(x,Y_{j-1})$ were encoded with the help of simple connected graphs with no cycles. More precisely, it was shown that the terms of leading contribution to (2.9) can be related with the set $T_k$ of one rooted trees with $k$ edges drawn in the upper half-plane. This representation is equivalent to the simple walks description of [14] and the number $|T_k|$ of elements in $T_k$ is given by $t_k$. Introducing some additional encoding for the trajectories $W$ that provide vanishing contribution, the authors have studied asymptotic behaviour of the moments $M_j^{(N)}$ for all $j \ll N^{1/6}, N \to \infty$ and proved boundedness of the limit of $\|\hat{A}_N\|$. These results were recently improved in paper [13], where the case of $j \ll N^{2/3}$ has been investigated.

Also let us note the paper [2], where the necessary and sufficient condition for finiteness of $\lim_{N\to\infty} \|A_N\|$ are found. These conditions are rather weak and require existence of the fourth moment of the random variables $a(x,y)$.

The methods developed in [2] and [13] are different from that used in [8]. It is not clear whether they are applicable to the dilute random matrices or not. In present paper we follow the way of [14] and [8] and develop it to be applied for the Wigner random matrices and their dilute versions as well.

There are three principal observations that we are based on:

(i) all paths $W(x,Y_{2k-1})$ with non-zero contribution can be separated uniquely into classes $\Pi(\tau)$ corresponding to the trees $\tau \in T_k$;

(ii) the paths that fall into the same class $\Pi(\tau)$ are described by graphs $\gamma(\tau)$ that are obtained from $\tau$ by gluing its vertices;

(iii) the contribution of the path is given by the number of cycles in $\gamma$.

These facts allow one to estimate easily the number of graphs $\gamma$ and to sum the corresponding contributions.

In Section 3 we give precise account on the graph representation of paths in the sum

$$M_j^{(N,p)} \equiv \mathbf{E}\langle N^{-1}\ \mathrm{Tr}\ [A_N^{(p)}]^j\rangle = \frac{1}{N}\sum_{x,y_i}\mathbf{E}A(x,Y_{2k-1})\langle D_N(x,Y_{2k-1})\rangle \tag{2.11}$$

where

$$A(x, Y_{j-1}) \equiv a(x, y_1)a(y_1; y_2) \cdots a(y_{j-1}, x)$$

and

$$D_N(x, Y_{2k-1}) \equiv \hat{d}_N(x, y_1) \cdots \hat{d}_N(y_{j-1}, x).$$

It follows from this graphical description that the terms of (2.11) that are of the order $O(p^{-s})$ are described by the graphs $\hat{\gamma}$ that have cycles only of the length 2. We show this in Section 4.

In Section 5 we estimate the number of trees that can produce the cycles of the length 2. Basing on relation (2.10), we develop a kind of the tree calculus and show that the number of trees having vertices of large degree is exponentially small with respect to the total number of trees.

In Section 6 we prove our main technical estimate

$$M_{2k}^{(N,p)} \leq t_k \sum_{r,s=1}^{k} \frac{(\alpha k)^{3r}}{N^r} \frac{(\beta k)^{s(1+\theta_k)}}{p^s} V_{4s} \tag{2.12}$$

with $\theta_k \to 0$ as $k \to \infty$. This inequality allows us to prove theorem 2.1. Theorem 2.2 is proved in Section 7.

We summarize our results and compare them with certain known facts from probability theory and related fields in Section 8. However, let us briefly discuss here one consequence of our results with respect to the Wigner random matrices that correspond to the case of $p = N$ in (1.5). It follows from (2.12) with $p = N$ that one can assume that moments $\mathbf{E}|a(x, y)|^{2m}$ exists for $m \leq m_0$ and obtain boundedness of $M_{2k}^{(N)}$ for all $k \ll N^\alpha$, where $\alpha$ depends on $m_0$. For example, using a version of standard truncation procedure (see e.g. a version presented in [2]), one can easily show that that $\alpha = 1/4$ requires $m_0 = 8$. This indicates a bridge between results of [2] and [13]. Relation (2.12) explains why the higher moments of the matrix entries $a(x, y)$ disappear from the limiting expressions for $M_{2k}$.

# 3 Even partitions and trees

It is clear that conditions (2.1a) imply that the average $\mathbf{E}A(x, Y_{j-1})$ is non-zero if and only if the path $W(x, Y_{j-1})$ is even [13], i.e. if each step $(y_i, y_{i+1})$ counted both in direct and inverse directions appears in $W$ even number of times. This immediately lead to the conclusion that $M_{2k+1}^{(N,p)} = 0$.

In the case when $j = 2k$, an even path uniquely determines a partition $\pi$ of variables $\{y_i, i = 0, \ldots, 2k-1\}$ into groups. We call such a partition the even one. The following statement links even partitions and rooted trees drawn in the upper half-plane. Let us note that given such a tree, one can order its edges. We adopt that the edges are enumerated from below and from the left. In other words, we say that the edge $e$ is less than the edge $e'$ when $e$ is situated on the

path from the root to $e'$ or when there is a vertex $\nu$ such that the path from $\nu$ to $e$ is to the left with respect to the path from $\nu$ to $e'$.

**Proposition 3.1.**
*The set $\Pi'_{2k}$ of all even partitions of the variables $(y_0, y_1, \dots y_{2k-1})$ can be uniquely separated into classes $\Pi(\tau)$ labeled by the trees $\tau \in T_k$.*

*Proof.* We are going to show that given a partition $\pi \in \Pi'_{2k}$, one can create a graph $g(\pi)$ that is uniquely determined by a tree $\tau$. We construct $g(\pi)$ with the help of natural procedure that simplifies approach suggested in [8] and resembles certain arguments of [13].

The first step is to determine the number of different groups of variables in $\pi$. This coincides with the number of vertices in $g(\pi)$. We refer to the vertex that corresponds to the group containing $y_0$ as to *the root.* Edges of $g(\pi)$ correspond to steps $s_i \equiv (y_i, y_{i+1})$, $i = 0, \dots, 2k-1$.

Next, one starts to go along $W = (y_0, \dots, y_{2k-1}, y_0)$ and draws the edges of $g$ starting from the root and joining subsequently the vertices according to the order dictated by appearance of steps $s_i$. One obtains closed connected graph with directed and ordered edges.

Let us consider two vertices joined by one or more edges. Since all edges of $g$ are ordered, the edges between two vertices are also ordered and can be numerated between themselves. We call the edge *marked* if it has an odd number in this inner enumeration. It is not hard to see that there are exactly $k$ marked edges in $g$.

Finally one has to go once more along $g$ and enumerate the marked edges according to the order of their appearance in $g$. Regarding the marked edges only, one obtains the connected graph $\gamma$ with $k$ numerated edges and one root. Let us call it *the first part* of the walk $g$. The unmarked steps of $g$ make also connected graph with $k$ edges that we call *the second part* of $g$.

The graph $\gamma$ has one root and $k$ numerated edges. One can uniquely restore the tree $\tau$ excepting the cases when the edges $i$ and $i+1$ in $\gamma$ have both vertices common. In this case we accept that $\tau$ is such that these vertices are consecutive.

Proposition is proved.□

Let us note in conclusion that graph $\gamma(\tau)$ is obtained from $\tau$ by gluing $q \geq 0$ vertices among them. One can regard this as if $\gamma$ obtained by a procedure of $q$ steps where on each step just one pair of vertices is glued.

Also it should be stressed that there exist partitions $\pi \neq \pi'$ that lead to the same graph $\gamma(\tau)$. The difference between $\pi$ and $\pi'$ is that they may have different second parts of $g$ and $g'$. This happens only when in $\gamma$ there are cycles of the length more than 2 with edges "correctly" enumerated, i.e. one can go along the the cycle and see edges with increasing numbers prescribed for the edges of $\tau$. We call these cycles as *the correct loops.* In all other cases the second part of $g$ is uniquely reconstructed from $\gamma$. We give precise account on this property in Section 4.

## 4 Number and contributions of partitions

It is obvious that the value of $\mathbf{E}A(x, Y_{2k-1})$ as well as that of $\langle D_N(x, Y_{2k-1})\rangle$ does not change if one preserves the partition and moves variables $y_i$. Therefore one can write relation

$$M_{2k}^{(N,p)} = \sum_{\pi \in \Pi'_k} B(\pi), \quad B(\pi) \equiv \frac{1}{N} \sum_{x,y_i}^{(\pi)} \mathbf{E}A(x, Y_{2k-1}) \langle D_N(x, Y_{2k-1})\rangle \tag{4.1}$$

and $\Sigma_{x,y_i}^{(\pi)}$ denotes the sum taken over variables $x, y_1, \dots, y_{2k-1}$ in the way such that the partition $\pi$ is preserved.

Tree representation helps to compute the contribution provided by the sum over given even partition $\Sigma^{(\pi)}$. This contribution depends on the form of the graph $\gamma(\tau)$ that corresponds to $\pi$.

Let us introduce several terms. Assume that in a tree $\tau$ there is a vertex $\nu$ of degree $m \geq 2$. The set of edges adjacent to $\nu$ we call *the cluster* of *the power* $m$. If two vertices of the edges belonging to the same cluster are glued, then we say that there is *the cluster gluing* in the tree $\tau$. The other gluings are called *the ordinary* ones.

**Proposition 4.1**.
*Assume that $\gamma(\tau)$ is obtained from $\tau$ with the help of $q = r+s$ gluings, where $s$ is the number of cluster gluings. Then*

$$B(\pi) = B(\gamma) \leq \frac{1}{N^r\, p^s} \prod_{i=1}^{j} V_{4\xi_i}. \tag{4.2}$$

*Remark.* We say that the graph $\gamma$ with $r+s$ gluings and corresponding partition $\pi$ are *of the type* $(r, s)$.

The proof of Proposition 4.1 can be easily derived from two observations.

The first one is that the number of vertices in $\gamma$ is equal to the number of groups of variables in $\pi$. Thus there are $(N-1)(N-2)\cdots(N-(k-q+1))$ terms in the sum $\Sigma^{(\pi)}$ of (4.1).

The second observation follows from the definitions (1.4) and (1.5). It is easy to see that if in $\gamma$ there exist two vertices $\mu$ and $\nu$ such that there are $\xi > 1$ edges $(\mu, \nu)$, then such a *multiple junction* provides the factor $\mathbf{E}a^{2\xi}\langle \hat{d}_N^{2\xi}\rangle = V_{2\xi} N^{-1} p^{1-\xi}$. If $\xi = 1$, then the factor due to this simple junction is obviously $1/N$.

It is clear that if there are $s$ gluings in one cluster, then one obtains $l$ multiple junctions with multiplicities $\xi_1, \dots, \xi_l$ such that $\xi_1 + \dots \xi_l = s + l$. Taking into account elementary inequalities

$$V_{2\xi_1} \cdots V_{2\xi_l} \leq V_{2\xi_1 + \dots 2\xi_l} = V_{2(s+l)} \leq V_{4s},$$

one arrives at (4.2). □.

Proposition 4.1 shows that all the partitions of the type $(r, s)$ provide the same contribution. Now it remains to estimate their number. This number can be estimated by the number $\mathcal{N}(r, s)$ of possibilities to make $r + s$ gluings in the tree $\tau$. We also should multiply $\mathcal{N}(r, s)$ by the number of different partitions that correspond to the same graph $\gamma$ of the type $(r, s)$.

Let us note here that the cluster gluings are independent from the other gluings in the sense that the number of possibilities to make the cluster gluing in $\gamma$ does not depend on the number of other gluings already performed. This observation together with (4.2) shows that one can treat these two types of gluings separately.

It is not hard to see that given a graph $\hat{\gamma}(\tau)$ of type $(o, s)$, corresponding partition $\hat{\pi}$ can be uniquely restored. This is because the way along $\hat{\gamma}$ preserving the existent order can be performed uniquely. In other words, the first and the second parts of the graph $\hat{g}$ are uniquely determined by $\hat{\gamma}$. Thus, the number of partitions of the type $(0, s)$ coincides with the number $D_s(\tau)$ of possibilities to make $s$ cluster gluings in $\tau$.

The picture differs for the ordinary gluings leading to the cycles of the length greater than 2. Assume that there is a vertex $\nu$ with $m > 3$ edges $e_i$ and there is a cycle of the length $l \geq 3$ starting and ending at $\nu$. If this is the correct loop, then one can reconstruct the second part of the graph $g$ in several ways because one can pass the cycle for the second time between passing the edges $e_i$. Also each loop can be passed in two opposite directions and this makes double the total number of possible ways.

Now let us estimate the number of partitions of the type $(r, s)$ performed in a tree $\tau$ that has $l$ clusters of powers $m_1, \ldots, m_l$, respectively. Clearly, $k \leq m_1 + \ldots m_l \leq 2k - 2$. Let us assume that there are $s$ cluster gluings such that in cluster $i$ there are $s_i$ gluings. Then, summarizing the arguments presented above, one can write that

$$M_{2k}^{(N,p)} \leq \sum_{\tau \in T_k} \sum_{q=0}^{k} \sum_{r+s=q} \frac{1}{N^r} \left[ D_\tau(r) \sum_{\{R_l\}}^{r} \mathcal{N}_\tau(\{R_l\}) \right] \frac{1}{p^s} \left[ \sum_{\{S_l\}}^{s} D_\tau(\{S_l\}) \prod_{i=1}^{l} V_{4s_i} \right]. \tag{4.3}$$

In this formula $D_\tau(r)$ is the number of possibilities to make $r$ ordinary gluings in the tree $\tau$, $\mathcal{N}_\tau(\{R_l\})$ is the number of different partitions that can be obtained from the graph $\gamma(\tau)$ due to presence of correct loops, and $D_\tau(\{S_l\})$ is the number of possibilities to make $s$ cluster gluings in $\tau$. Here we have denoted $R_l = (r_1, r_2, \ldots, r_l)$, and $S_l = (s_1, s_2, \ldots, s_l)$, such that $r_1 + r_2 + \ldots + r_l = r$. Obviously, $s_1 + s_2 + \ldots + s_l = s$. Summations go over all possible combinations of $r_i \geq 0$ and $s_i \geq 0$ satisfying conditions presented above.

Elementary calculation shows that

$$D_\tau(r) \le \frac{1}{r!}\binom{k}{2}\binom{k-1}{2}\cdots\binom{k-r+1}{2} \le \frac{k^{2r}}{2^r r!}. \qquad (4.4)$$

Next, assuming that all cycles obtained are correct, we can write that

$$\mathcal{N}_\tau(r) \equiv \sum_{\{R_l\}}^{r} \mathcal{N}_\tau(\{R_l\}) \le \sum_{\{R_l\}}^{r} \prod_{i=1}^{l} \frac{(m_i-2)!}{(m_i-1-r_i)!}. \qquad (4.5)$$

One can obtain the latter expression regarding the observation that if in a cluster $i$ there are $r_i$ correct loops, then the number of different ways to pass them is estimated by the number of possibilities to distribute $r_i$ different balls into $m_i - r_i$ boxes. We easily derive from (4.5) that

$$\mathcal{N}_\tau(r) \le \sum_{\{R_l\}}^{r} \prod_{i=1}^{l} 2^{r_i} m_i^{r_i} \le 2^r \sum_{\{R_l\}}^{r} r! \prod_{i=1}^{l} \frac{m_i^{r_i}}{r_i!} \le (m_1 + \ldots + m_l)^r = (4k)^r. \quad (4.6)$$

Let us turn to the number of cluster gluings. It is estimated by the product

$$\prod_{j=1}^{s} \frac{1}{s_j!}\binom{m_j}{2}\binom{m_j-1}{2}\cdots\binom{m_j-s_j+1}{2} \le \frac{1}{s!}\frac{s!}{s_1!\cdots s_j!}\prod_{j=1}^{s} m_j^{2s_j}.$$

Taking into account inequality $V_{4s_1}\cdots V_{4s_l} \le V_{4s}$, we can write that

$$\sum_{\{S_l\}}^{s} D_\tau(\{S_l\}) \prod_{i=1}^{l} V_{4s_i} \le \frac{(m_1^2 + \ldots + m_l^2)^s}{s!} V_{4s}. \qquad (4.7)$$

Trivial inequality for the variable $\Sigma_k^{(2)}(\tau) \equiv m_1^2 + \ldots m_l^2 \le 4k^2$ lead to the estimate $k^{2s}/p^s$. This estimate is rather rough and is not sufficient for the proof of theorem 2.1. In the next section we show that $\sum_\tau \Sigma_k^{(2)}(\tau)$ behaves like $t_k k^{1+\delta}$ that reflects the almost linear character of the average tree. This fact together with inequalities (4.5)-(4.7) implies (2.12).

# 5 Enumeration of trees

In this section we consider the set $T_k$ the one-root trees $\tau$ having $k$ edges drawn in the upper half-plane. Let us recall that given a tree $\tau \in T_k$, one can order its edges. Let us also give several definitions.

We refer to the edges adjacent to the root as to the *root edges.* We define the *cluster* $\kappa$ as the set of $m$ edges that have one common vertex $\nu$. We call $m$ and $\nu$ the *power* and the center, respectively, of the cluster $\kappa$.

The main result of this section is given by the following statement.

**Theorem 5.1.**
*Let us consider the set $G_k^{(m)} \subset T_k$ of trees that have at least one cluster with the power $m \geq 2$. Then the number of such trees $|G_k^{(m)}| \equiv g_k(m)$ satisfies relation*

$$g_k(m) \leq k t_k \left(\frac{3}{4}\right)^{m-2}. \tag{5.1}$$

To prove theorem 5.1, we need the following auxiliary statement.

**Lemma 5.1.**
*Let us consider the set $T_k^{(m)}$ of trees that have $m \geq 2$ root edges. Then*

$$t_k^{(m)} \equiv |T_k^{(m)}| \leq t_{k-1} \left(\frac{3}{4}\right)^{m-2}. \tag{5.2}$$

*Proof.* It is easy to see that

$$t_k^{(m)} = \sum_{\alpha_i}^{k-m} t_{\alpha_1} t_{\alpha_2} \cdots t_{\alpha_m}, \tag{5.3}$$

where the sum is taken over all $\alpha_i \geq 0$ such that $\alpha_1 + \ldots + \alpha_m = k - m$.

Let us derive first the simple estimate

$$t_k^{(m)} \leq t_{k-1}. \tag{5.4}$$

Regarding (2.10), we can rewrite (5.3) in the form

$$t_k^{(m)} = \sum_{q=0}^{k-m} \sum_{\alpha_i}^{k-m-q} t_{\alpha_1} t_{\alpha_2} \cdots t_{\alpha_{m-2}} \sum_{\alpha_i}^{q} t_{\alpha_{m-1}} t_{\alpha_m} =$$

$$\sum_{q=0}^{k-m} \sum_{\alpha_i}^{k-m-q} t_{\alpha_1} t_{\alpha_2} \cdots t_{\alpha_{m-2}} t_{q+1}. \tag{5.5}$$

The latter sum can be regarded as the sum over $m-1$ variables $\alpha_i \geq 0$, where the term with $\alpha_{m-1} = 0$ is absent. Then

$$t_k^{(m)} = \sum_{\alpha_i}^{k-m+1} t_{\alpha_1} t_{\alpha_2} \cdots t_{\alpha_{m-1}} - \sum_{\alpha_i}^{k-m+1} t_{\alpha_1} t_{\alpha_2} \cdots t_{\alpha_{m-2}} \equiv t_k^{(m-1)} - t_k^{(m-2)}.$$

Thus, $t_k^{(m)} \leq t_k^{(2)}$. Relation (2.10) implies equality $t_k^{(2)} = t_{k-1}$ that gives (4.4).

Let us rewrite (5.3) in the form

$$t_k^{(m)} = \sum_{q=0}^{k-m} t_{k-q-1}^{(m-1)} t_q. \tag{5.6}$$

If $m-1 \geq 2$, then we can apply (4.4) to $t_{k-q-1}^{(m-1)}$ and obtain that

$$t_k^{(m)} \leq \sum_{q=0}^{k-m} t_{k-q-2} t_q =$$

$$t_{k-1} - (t_0 t_{k-2} + \ldots + t_{m-1} t_{k-m-1}) \leq t_{k-1} - t_{k-2}.$$

Expression (2.8) for $t_k$ implies that $t_{k-2} > t_{k-1}/4$. Therefore $t_k^{(m)} \leq 3t_{k-1}/4$.

If $m-2$ is greater than 2, then we can substitute the last inequality into (5.6) and obtain that $t_k^{(m)} \leq t_{k-1}(3/4)^2$. Now it is clear that (5.2) is true.

Using similar computations, one can easily prove the following statement.

**Lemma 5.2**.

*Let us denote by $n_r(s)$ the number of trees that can be constructed on $r \geq 1$ roots with the help of $s$ edges. Then for $2 \leq l \leq s$*

$$n_{r+l}(s-l) \leq n_{r+1}(s-1)\left(\frac{3}{4}\right)^{l-2} \leq n_r(s)\left(\frac{3}{4}\right)^{l-1}. \tag{5.7}$$

*Proof theorem 5.1.*

We start with the observation that one can construct a tree $\tau \in T_k$ from the set $E_k$ of $k$ edges $e_i$ that are already enumerated on the way that this enumeration will agree with the order among tree edges.

Now assume that before this procedure the edges $e_h, e_{h+1}, \ldots, e_{h+m}$ are joined to the same vertex and used in the construction as one cluster. According to Lemma 4.2, this diminishes exponentially with respect to $t_k$ the number of trees obtained on this way. Taking into account that $h$ can vary from 1 to $k-m$ we obtain the estimate (5.1).$\square$

In conclusion let us note that (5.1) implies that the number of trees that have the power of the maximal cluster large enough is exponentially small with respect to $t_k$. Therefore the number of such clusters is exponentially small with the total number of clusters in trees $\tau \in T_k$. This means that the average degree of a tree vertex remains finite even in the limit $k \to \infty$. We plan to study this problem mores systematically in a separate publication.

## 6 Proof of theorem 2.1.

As it follows from (5.3) and (5.7), we need to estimate the sum

$$Q_k(s) \equiv \sum_{s=1}^{k-1} \frac{V_{4s}}{s! p^s} \sum_{\tau \in T_k} (m_1^2 + \ldots + m_l^2)^s,$$

where $m_i$ are the powers of clusters of $\tau$. Let us denote by $\hat{m}(\tau)$ the maximal cluster power of the tree $\tau$; then obviously

$$\sum_{\tau \in T_k} (m_1^2 + \ldots + m_l^2)^s \le \sum_{\tau \in T_k} \hat{m}(\tau)^s (m_1 + \ldots + m_l)^s = (2k)^s \sum_{\tau \in T_k} \hat{m}(\tau)^s.$$

Let us consider the set of trees $T_k'$ such that $\hat{m}_\tau \le \chi \log k$, where $\chi = (\log 4 - \log 3)^{-1}$. Then

$$(2k)^s \sum_{\tau \in T_k'} \hat{m}_\tau^s \le t_k (2k\chi \log k)^s \le t_k (2\chi k)^{s(1+\theta)}$$

with some $\theta = \delta_k$ that can be taken vanishing as $k \to \infty$.

We represent the set $T_k \setminus T_k'$ as the sum of sets $G_k([\chi(j + \log k)])$ of trees that have the maximal cluster power equal to $[\chi(j + \log k)]$, $j = 1, 2, \ldots$. Here we denoted by $[x]$ the maximal natural number less than or equal to $x$. Using estimate (5.1), we can write that

$$\sum_{\tau \in T_k'} \hat{m}_\tau^s \le \sum_{j=1}^{k-[\chi \log k]} \sum_{\tau \in G_k(j+[\chi \log k])} \hat{m}_\tau^s \le$$

$$e^2 \sum_{j=1}^{k} \chi^s (\log k + j)^s \exp\{-j - \log k\} k t_k \le e^2 (s+1)! (\chi \log k)^s.$$

Thus, we obtain inequality (cf. (2.12))

$$Q_k(s) \le p^{-s} (\beta k)^{s(1+\theta')} V_{4s}, \tag{6.1}$$

where $\beta$ is a constant and $\theta'$ can be chosen vanishing when $k \to \infty$.

Now it is easy to derive the estimate (2.4) from (2.12). Taking into account (6.1) and the estimate $V_{4k} \le (2k)^{2\delta k}$, we obtain inequality

$$M_{2k}^{(N,p)} \le t_k \left( (1 - \alpha k^3 N^{-1}) (\beta k^{1+2\delta+\theta'} p^{-1}) \right)^{-1}.$$

Then we can deduce that inequality

$$M_{2k}^{(N,p)} \le \varepsilon t_k \le (1+\varepsilon)^{2k} \, 4^k \tag{6.2}$$

holds for all $k$ sufficiently large and such that $k \ll N^{1/3}$ and $k^{1+\delta+\theta'} \ll p$.

Regarding definition (1.2), we can write that

$$M_{2k}^{(N,p)} \leq \mathbf{E}\langle \int_{|\lambda|\geq 2(1+2\varepsilon)} \lambda^{2k} \mathrm{d}\sigma(\lambda; A_N^{(p)})\rangle \geq 4^k (1+2\varepsilon)^{2k} N \mathbf{E}\langle \#\{|\lambda_j^{(N,p)}| \geq 2(1+\varepsilon)\}\rangle \geq$$

$$4^k (1+2\varepsilon)^{2k} N \mu_a \otimes \mu_d \{\omega : \|A_N^{(p)}\| > 2(1+2\varepsilon)\}.$$

Combining these inequalities with (6.2), we derive that

$$\mu_a \otimes \mu_d \{\omega : \|A_N^{(p)}\| > 2(1+2\varepsilon)\} \leq N \left(\frac{1+\varepsilon}{1+2\varepsilon}\right)^{2k}.$$

Choosing $k$ such that $k/\log N \to \infty$, we obtain that (2.4) holds. Theorem 2.1 is proved.

# 7 Proof of theorem 2.2

Let us consider the unit vectors $\vec{e}^{(j)}$ with the components

$$\vec{e}^{(j)}(x) = \begin{cases} 1, & \text{if } x = j, \\ 0, & \text{if } x \neq j. \end{cases}$$

We can write that

$$\|A_{N,p}\|^2 \geq \max{}_{j=1,\ldots,N} \left\| A_{N,p} \vec{e}^{(j)} \right\|^2 = \max{}_{j=1,\ldots,N} \left(A_{N,p}^2\right)(j,j), \qquad (7.1)$$

where

$$\left(A_{N,p}^2\right)(j,j) = \sum_{y=1}^{N} |A_{N,p}(j,y)|^2 .$$

Let us introduce random variables

$$h_j^{(N)} = \sum_{y=j}^{N} |A_{N,p}(j,y)|^2 ,\ j = 1, \ldots, N.$$

It is clear that the family $\left\{h_j^{(N)}\right\}_{j=1}^{N}$ is the set of jointly independent random variables and that

$$\max_{j=1,\ldots,N} \left(A_{N,p}^2\right)(j,j) \geq \max_{j=1,\ldots,N} h_j^{(N)}. \qquad (7.2)$$

Let us note that under conditions of theorem 2.2

$$h_j^{(N)} = \frac{1}{p} \sum_{y=j}^{N} d(j,y) \equiv \frac{1}{p} \eta_j^{(N)}. \qquad (7.3)$$

Let us consider the probability distribution of the random variable

$$H^{(N)} = \max_{j \le N/2} h_j^{(N)}.$$

It is clear that

$$P_N(R) \equiv \Pr\{H_N < R\} = \Pr\left\{\max_{j \le N/2} \eta_j^{(N)} < pR\right\}$$

$$= \prod_{j=1}^{N/2} \left(1 - \Pr\left\{\eta_j^{(N)} \ge pR\right\}\right).$$

It is not hard to prove that the random variables $\eta_j^{(N)}$ considered for $p \sim \log N$ converge as $N \to \infty$ to the Poissonian random variables $\zeta_j$ that are identically distributed with the parameter $p$. Then we can write that

$$1 - \Pr\left\{\eta_j^{(N)} \ge pR\right\} \le 1 - \Pr\{\zeta_j = pR\} = 1 - \frac{p^{pR}}{(pR)!} e^{-p}.$$

Let us show that if $p = (\log\ N)^{1-\delta}$, then $p^{pR}\left[e^p (pR)!\right]^{-1}$ decays more slowly than $2/N$ and

$$\left[1 - \frac{p^{pR}}{(pR)!} e^{-p}\right]^{\frac{N}{2}} \to 0 \ as\ N \to \infty. \tag{7.4}$$

Using the Stirling formula, we can write that

$$e^{-p} \frac{p^{pR}}{(pR)!} \sim \frac{e^{p(R-1)}}{R^{pR}} = \exp\left\{-pR\left[\log R - \frac{r}{r-1}\right]\right\}.$$

It is easy to see that (7.4) holds that together with (7.1) and (7.2) proves relation (2.6).

# 8 Summary and discussion

We consider the ensemble of random matrices $A_{N,p}$ with independent entries such that $A_{N,p}$ has, in the average $p$ non-zero entries per row. We study asymptotic behaviour of the spectral norm $\|A_{N,p}\|$ in the limit when $N, p \to \infty$ and $p = o(N)$. We consider the averaged moments

$$M_{2k}^{(N,p)} = \frac{1}{N} \sum_{x=1}^{N} \sum_{\{y_i\}} \mathbf{E}\langle A_{N,p}(x, y_1) \cdots A_{N,p}(y_{2k-1}, x)\rangle \tag{8.1}$$

and give a further development to the approach proposed in papers [8, 14]. The method is based on the relation between the even partitions $\pi$ of variables

$(x, y_1, \ldots, y_{2k-1})$, i.e. those that make the average in (8.1) non-zero, from one hand and the rooted trees $\tau \in T_k$ constructed in the upper half-plane with the help of $k$ edges, from the other hand.

We give the full description for the even partitions $\pi$ in terms of the graphs $\gamma$ obtained from a tree $\tau$ by gluing its vertices. The magnitude of the terms described by partition $\pi$ is determined by the number of cycles in corresponding graph $\gamma$. This allows one to estimate easily the number of different partitions and their contributions to $M_{2k}^{(N,p)}$.

We show that the terms of the order $p^{-l}\ M_{2k}^{(N,p)}$ in the limit $p = o(N)$, $N \to \infty$ comes from those partitions $\pi'$ that are encoded by the graphs $\gamma'$ that have cycles of the length 2 exactly. This means that the average degree of the vertex plays the key role in asymptotic properties of large sparse random matrices.

We obtain the estimate showing that the average in certain sense tree is of the linear structure and the relative number of possible 2-cycles increases almost linearly with respect to $k \to \infty$. This allows us to derive our main result that $\|A_{N,p}\|$ remains bounded when $p \sim (\log N)^{1+\delta}$, $\delta > 0$. We show that the value $p_c = \log N$ is the critical one because $\|A_{N,p}\|$ is unbounded when $p \sim (\log N)^{1-\delta}$.

Our results can be compared with the classical Erdös-Rényi limit theorem [6]. This statement concerns independent random variables $\xi_1, \ldots, \xi_N$ and the partial sums $X_{i,p} = \xi_i + \xi_{i+1} + \ldots + \xi_{i+p}$. It is proved that the value $p_c = \log N$ is the critical one for the $X(N) = \max_i X_{i,p} p^{-1}$ to be either bounded or not.

One can regard our theorems 2.1 and 2.2 as a limit theorems for stochastic versions of Erdös-Rényi partial sums. This claim is supported by the fact that random variables $h_j^{(N)}$ (5.3) are determined as the sum of approximately $p$ random variables. Let us also note that theorem 2.1 concerns a random variable $\|A_{N,p}\|$ that is greater than $\hat{X}(N) \equiv \max_j \|A_{N,p}\vec{e}^{(j)}\|$. This could explain the need of conditions (2.2) that are more restrictive than the standard conditions of Erdös-Rényi limit theorem.

One can also trace out more subtle link between our results and the well-known statement concerning the connectedness of random graphs (see e.g.[3]). If one has $N$ vertices and draws $q$ edges joining randomly chosen vertices, then the graph obtained will be connected in the limit $N \to \infty$ provided $q \gg N \log N$. In our terms the randomly chosen pairs $(i, j)$ to be joined correspond to non-zero elements in the dilute matrix $A_{N,p}$. In this context, the transition from disconnected graph to the connected one can be regarded as analog of the transition of $A_{N,p}$ from the class of the tridiagonal matrices (like discrete Schrödinger operator) to the class of Wigner random matrices.

It should be noted that in mathematical physics literature another critical values are found for sparse random matrices. In particular, in paper [10] it is claimed that certain "density-density" correlation function changes its behaviour at finite values of $p$. Numerical studies [7] show that certain spectral characteristics of strongly dilute random matrices can depend on finite values of

$p$. Therefore our results imply that there can be several different critical values in the sparse random matrix model.

To complete the discussion, let us note that the dilute random matrices $A_{N,p}$ with $p$ replaced by $N$ take the form of the Wigner random matrices $\hat{A}_N$. Therefore technique developed in present paper can be also useful in this case. This topic is already addressed in Section 2.

**Acknowledgments.** The author is grateful to Dr. V.Gayrard and Dr. A.Soshnikov for useful discussions and to Prof. F.Comets for his comments about the Erdös-Rényi limit theorem. The author also thanks Laboratory of Mathematical Physics and Geometry, University Paris-7, where the paper is completed, for kind hospitality and Prof. A.Boutet de Monvel for her interest to this work.

# References

[1] D.J. Amit. *Modeling Brain Function.* Cambridge University Press (1989)

[2] Z.D. Bai, Y.Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.* **16** (1988) 1729-1741

[3] B. Bollobas. *Random Graphs.* Academic Press, New York (1985)

[4] A. Bovier and V. Gayrard. Rigorous results on the thermodynamics of the dilute Hopfield model. *J. Stat. Phys.* **72** (1993) 79-112

[5] A. Crisanti, G. Paladin, and A. Vulpiani, *Products of Random Matrices in Statistical Physics.* Springer Series in Solid State Sciences, vol. 104. (1993) Springer, Berlin, New York.

[6] P. Erdös and A. Rényi. On a new law of large numbers. *J. Analyse Math.* **23** (1970) 103-111

[7] S. Evangelou. A numerical study of sparse random matrices. *J.Stat. Phys.* **69** (1992) 361

[8] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica* , **1** (1981) 233-241

[9] A. Khorunzhy, B. Khoruzhenko, L. Pastur , and M. Shcherbina Large-$n$ limit instatistical mechanics and the spectral theory of disordered systems. In: *Phase Transitions and Critical Phenomena*, eds. C. Domb and J.L. Lebowitz Vol. 15 (1992) (London: Academic Press), pp. 73-245

[10] A.D. Mirlin and Y.V. Fyodorov. Universality of level correlation fucntion of sparse random matrices. *J. Phys. A: Math. Gen.* **24** (1991) 2273

[11] C.E. Porter (ed.) *Statistical Theories of Spectra: Fluctuations* Academic Press, New York (1965)

[12] G.J. Rodgers and A.J. Bray. Density of states of sparse random matrix. *Phys. Rev. B* **37** (1988) 3557-3562

[13] Ya.G. Sinai and A.B. Soshnikov. A refinement of Wigner's semicircle law in a neighborhood of the spectrum edge for random symmetric matrices. *Func. Anal. and Its Appl.* **32** (1998) 114-131

[14] E. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Ann. Math.* **62** (1955) 548-564